Supplementary materials and methods

Clinicopathological information in breast cancer patients

Estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67 expression status were available from pathological reports. ER and PR were defined as positive with more than 10% positively staining nucleus [1]. Fluorescence *in situ* hybridization was performed to determine the HER2 status in the equivocal evaluation of HER2 expression by immunohistochemistry (IHC) [2]. The Ki-67 index was dichotomized to high- and low-expression groups, with 20% as the cut-off point [3]. Patients were staged using the American Joint Committee on Cancer (AJCC) for breast cancer staging [4]. All specimens were evaluated by two independent pathologists. Patient age, menstrual status, tumor size, histological grade, and lymph node status were obtained from medical records and pathological reports.

Breast cancer molecular subtypes

Patients were further classified based on surrogate definitions of ER, PR, HER2, and Ki-67 according to St. Gallen Consensus 2013 [5]: luminal A: ER-positive and/or PR-positive, HER2-negative, and Ki-67 ≤ 20%; luminal B (HER2-negative): ER-positive and/or PR-positive, HER2-negative, and Ki-67 > 20%; luminal B (HER2-positive): ER-positive and/or PR-positive, and HER2-positive; HER2 type: ER-negative, PR-negative, and HER2-positive; Triple-negative breast cancer (TNBC): ER-negative, PR-negative, and HER2-negative.

Lentiviral construction and cell infection

The short hairpin RNAs (shRNAs) targeting human YEATS4 and ZEB1 were designed and synthesized by GenePharma. The pLKO.1 constructed with specific shRNA and negative control shRNA (shCtrl), pCMVΔynthe, and pMD.G plasmids were co-transfected in 293T cells to package and produce lentivirus. Breast cancer cell lines were infected with the collected viral supernatant. The shRNA sequences are listed in Table S1. The plasmids of pcDNA3.1-control and pcDNA3.1-YEATS4 were transfected into cells by instantaneous transfection using X-tremeGENE™ HP DNA Transfection Reagent (Roche). A full-length coding sequence (CDS) of YEATS4 was implanted into the lentiviral vector pHIV-EGFP (Addgene) with an EF1-α promoter, a multiple cloning site (MCS), and EGFP. The lentiviral supernatant was obtained by transfecting 293T cells with pHIV-EGFP-YEATS4, pMDLg/pRRE, pMD2.G, and pRSV-Rev.

Cell proliferation assays

Cell proliferation was examined using the Cell Counting Kit-8 (CCK-8; Dojindo). Cells were cultured in 96-well plates (Corning) at a concentration of 2000 cells/well in a volume of 200 μL after stable YEATS4 deletion. At the indicated time points, 20 μL CCK-8 reagents were added to each well and then incubated for 2 h at 37°C. The absorbance at 450 nm was measured with a microplate reader. Each independent experiment was repeated three times.

Colony formation and soft agar colony formation assays

Colony formation assays: Cells stably transfected with OSR1-shRNAs or stably OSR1-expressing cells were digested into single cells, which were cultured in a 6-well plate with a density of 1×10^3 cells per well with 5% CO_2 at 37°C for 2 weeks. After colony formation, the plate was gently rinsed with PBS, fixed with 4% paraformaldehyde for 20 min, and stained with 0.2% crystal violet for 10 min. Colonies composed of more than 50 cells were then counted for the colony-forming assays. Three independent assays were performed to confirm the results.

Soft agar colony formation assays: The tested cells (1×10^3) were grown in 1 mL of complete medium containing 0.3% low-melting-point agarose gel (Sigma-Aldrich) and rested on another layer of 1.5 mL of 0.6% agarose gel in 6-well plates. The cells were then incubated with 5% CO_2 at 37°C for 2 weeks. Colonies larger than 50 μm in diameter were observed and counted under the microscope. Three independent experiments were performed to confirm the results.

Wound healing assays

Cells were cultured in 6-well plates (Corning) to reach more than 90% confluency in a monolayer. The scratched wounds were made using a sterile 200 μL pipette tip in the cell monolayer. The wells were rinsed with PBS to remove cellular fragments and incubated with 5% CO_2 at 37°C. Images were then taken with an optical microscope. The cell migrating distance was measured by gauging the scratch space at 0 h, 24 h, and 48 h by using the ImageJ software. Each independent experiment was repeated three times.

Transwell assays

For the invasion assays, breast cancer cells (2×10^4 cells) in a serum-free medium were plated into transwell chambers (8 μm pore size; Corning) coated with 60 μL of Matrigel (1:8 dilution; BD Biosciences). For cell migration assays, no Matrigel was added. The culture medium containing 20% FBS was filled in the lower compartment as a chemical attractant. After incubation for 24-48 h, the cells in the upper chamber were removed with a cotton tip. The cells on the bottom side of the transwell membrane were fixed with 4% paraformaldehyde for 20 min and stained with 0.25% crystal violet staining solution for 30 min. The cells were photographed with an optical microscope. The number of cells passing through the membrane in five random fields was determined using the ImageJ software. Each independent experiment was repeated three times.

Western blot analysis

Cellular protein was extracted using RIPA buffer (Sigma-Aldrich). Protein concentrations were quantified using the Bicinchoninic Acid Kit (Sigma-Aldrich) as instructed by the manufacturer. Proteins were resolved using 10% SDS-PAGE gel electrophoresis. They were then transferred to a polyvinylidene fluoride membrane (Millipore), blocked in 5% non-fat milk (Sigma-Aldrich) for 2 h, and incubated with the primary antibody (anti-YEATS4, Sigma-Aldrich, diluted by 1:1000; anti-GAPDH, Cell Signaling Technology, diluted by 1:1000) at 4°C overnight. Membranes were then incubated with the HRP-linked secondary antibody (diluted by 1:2000; Cell Signaling Technology) for 1 h at room temperature. Blots were visualized by using an ECL Detection Kit (Millipore) and analyzed using the ImageJ software.

Immunofluorescence

Cells growing on glass cover slides were fixed with 4% paraformaldehyde for 15 min at room temperature, permeated with PBS containing 0.25% Triton X-100 for 30 min, blocked with 1% BSA for 1 h at room temperature, and incubated overnight with the indicated primary antibodies at 4°C and then with fluorescence-conjugated secondary antibodies. Nuclei were stained with DAPI. Cells were visualized by confocal laser scanning microscopy (Zeiss LSM 880).

Dual luciferase reporter assays

Luciferase reporter assays were conducted in accordance with the instructions provided by the manufacturer. The ZEB1 promoter was inserted into the pGL3-Basic luciferase expression vector (Genepharma). Cells were grown in 12-well plates (Corning) and transfected with an appropriate pcDNA3.1 plasmid, together with a luciferase plasmid, by using the X-tremeGENE HP DNA Transfection Reagent (Roche). After transfection for 48 h, cell lysates were obtained and luciferase reporter activities were measured using the Dual-Luciferase Reporter System (Promega). Firefly luciferase activities were normalized to the activities of Renilla luciferase as the control. Each experiment was repeated three times.

In vivo tumorigenicity and metastasis assays

Female athymic nude mice (BALB/c-nu/nu) aged 4-6 weeks were obtained from Shanghai SLAC Laboratory Animal Company (Shanghai, China) and kept in specific-pathogen-free (SPF) environments at the Experimental Animal Center of Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine. For tumorigenicity assays, tumor cells were resuspended in phosphate-buffered saline (PBS)

at a concentration of 2×10^6/100 μL and injected subcutaneously into the nude mice (n = 6 for each group). The xenograft tumor size was evaluated every 3 d with a Vernier caliper. Tumor volume was calculated using the following formula: volume (mm^3) = [(shortest length)^2 × longest length]/2. Growth curves were plotted and analyzed for every group. All mice were euthanized after 40 d to isolate and collect tumors. The tumor weight was measured and recorded. For lung metastasis assays, tumor cells were resuspended in PBS at a density of 2×10^6/100 μL and intravenously injected into the nude mice (n = 8 for each group). The survival of all mice was recorded throughout the experiment. After 9 weeks, all mice were euthanized, and their lungs were resected for standard histological examinations.

References

[1] Yip CH and Rhodes A. Estrogen and progesterone receptors in breast cancer. Future Oncol 2014; 10: 2293-2301.

[2] Elebro K, Bendahl PO, Jernström H and Borgquist S. Androgen receptor expression and breast cancer mortality in a population-based prospective cohort. Breast Cancer Res Treat 2017; 165: 645-657.

[3] Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B and Senn HJ. Tailoring therapies--improving the management of early breast cancer: St gallen international expert consensus on the primary therapy of early breast cancer 2015. Ann Oncol 2015; 26: 1533-1546.

[4] Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ and Hortobagyi GN. Breast cancer-major changes in the American joint committee on cancer eighth edition cancer staging manual. CA Cancer J Clin 2017; 67: 290-303.

[5] Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B and Senn HJ. Personalizing the treatment of women with early breast cancer: highlights of the St gallen international expert consensus on the primary therapy of early breast cancer 2013. Ann Oncol 2013; 24: 2206-2223.

Table S1. ShRNA and primer sequences used in the research

Application	Gene	Sequence
Small hairpin RNA (shRNA)	YEATS4 #1	GCGGGAGAGTAAAGGGTGTTA
		TAACACCCTTTACTCTCCCGC
	YEATS4 #2	GGGAGAGTAAAGGGTGTTACT
		AGTAACACCCTTTACTCTCCC
	ZEB1	GGAGGATAAAGAGATGGAAGA
		TCTTCCATCTCTTTATCCTCC
Real time quantitative polymerase chain reaction (RT-qPCR)	YEATS4	GAGAATGGCCGAATTTGGGC
		CCGAGCAACATTACCGTAAACT
	Snail	ACTGCGACAAGGAGTACACC
		GAGTGCGTTTGCAGATGGG
	Slug	CGAACTGGACACACATACAGTG
		CTGAGGATCTCTGGTTGTGGT
	Twist1	GTCCGCAGTCTTACGAGGAG
		GCTTGAGGGTCTGAATCTTGCT
	ZEB1	GATGATGAATGCGAGTCAGATGC
		ACAGCAGTGTCTTGTTGTTGT
	FOXC2	CCTCCTGGTATCTCAACCACA
		GAGGGTCGAGTTCTCAATCCC
	EZH2	AATCAGAGTACATGCGACTGAGA
		GCTGTATCCTTCGCTGTTTCC
	BMI1	CGTGTATTGTTCGTTACCTGGA
		TTCAGTAGTGGTCTGGTCTTGT
	E-cadherin	CGAGAGCTACACGTTCACGG
		GGGTGTCGAGGGAAAAATAGG
	ZO-1	CAACATACAGTGACGCTTCACA
		CACTATTGACGTTTCCCCACTC
	Vimentin	GACGCCATCAACACCGAGTT
		CTTTGTCGTTGGTTAGCTGGT
	N-cadherin	TTTGATGGAGGTCTCCTAACACC
		ACGTTTAACACGTTGGAAATGTG
	E47	ACGAGCGTATGGGCTACCA
		GTTATTGCTTGAGTGATCCGGG
	GAPDH	TGTGGGCATCAATGGATTTGG
		ACACCATGTATTCCGGGTCAAT
ChIP assays	ZEB1 site A	TATTCGAAGGAGGTGGGAAGCAGG
		CGTGCAGGACCTTAAGGCAAGAAG
	ZEB1 site B	ACCGCTGATGAACTTTCCCA
		TGGAACTACCACAACGAGGC
	ZEB1 site C	CGTCCCCTAAGGCCAATACC
		AGCGTATGCATTTCGTTGCC

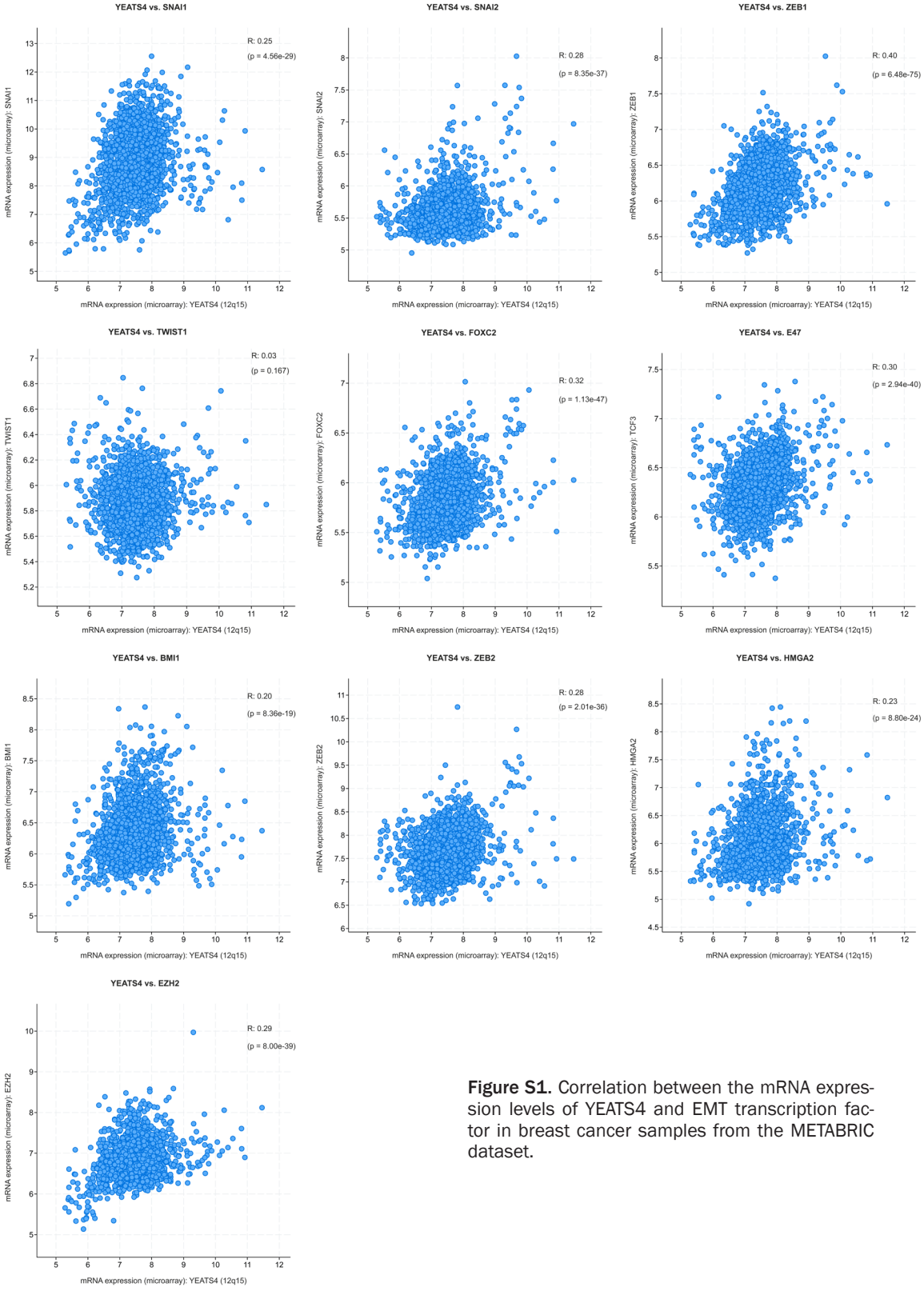

Figure S1. Correlation between the mRNA expression levels of YEATS4 and EMT transcription factor in breast cancer samples from the METABRIC dataset.

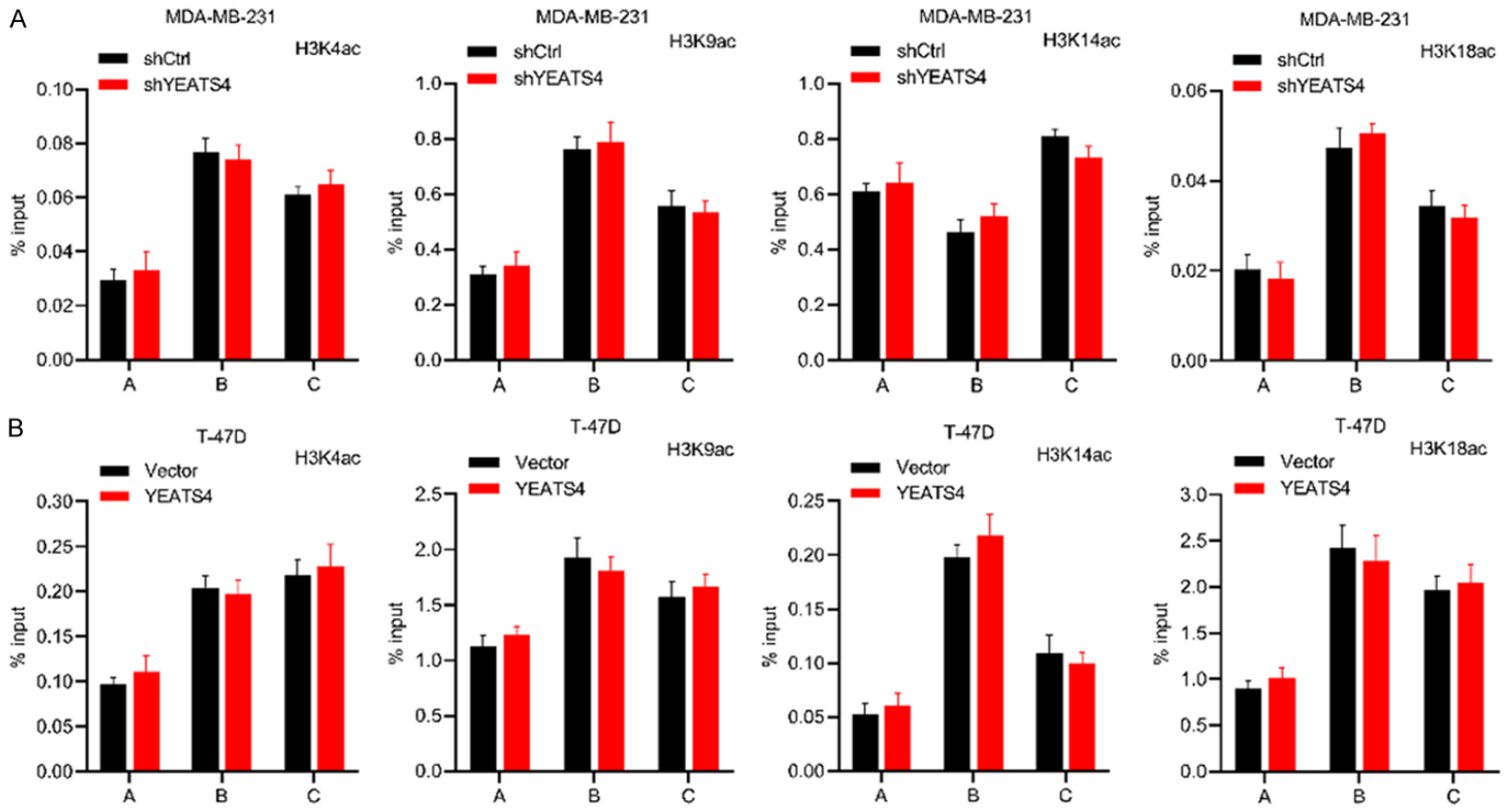

Figure S2. ChIP assays were performed to determine H3K4ac, H3K9ac, H3K14ac, and H3K18ac occupancy in MDA-MB-231-shYEATS4 (A) and T-47D-YEATS4 (B) cells. Data were presented as mean ± standard deviation for 3 independent assays.

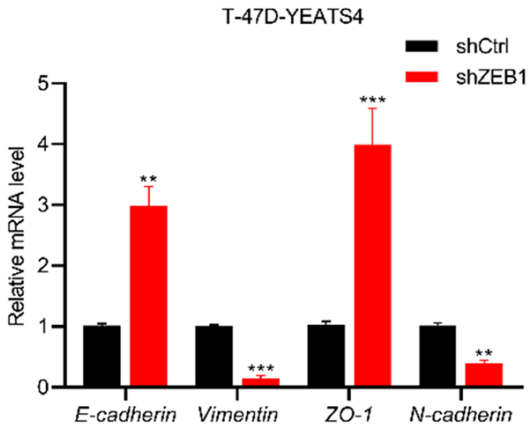

Figure S3. Deleting ZEB1 restored the expression of the epithelial markers E-cadherin and ZO-1 and reduced the mesenchymal markers vimentin and N-cadherin in T-47D-YEATS4 cells, as determined by real-time quantitative polymerase chain reaction (RT-qPCR). Data were presented as mean ± standard deviation for 3 independent assays. **, $P < 0.01$; ***, $P < 0.001$.